Integrating ENCODE data to interpret regulatory changes in cancer

Abstract

Cancer is caused by genomic variants that disrupt the normal physiology of living cells. Though mutations in coding genes have been well characterized, the preponderance of mutations in tumor genomes fall within non-coding regions, rendering them less amenable to characterization and interpretation. Newly released ENCODE data provides an opportunity to bridge these knowledge gaps. For a variety of cancer-derived cell lines, as well as non-cancerous cell lines derived from relevant tissues, ENCODE provides diverse genome-wide assays, such as Repli-seq, ChIP-seq, DNase-seq, STARR-seq, Hi-C, and ChIA-PET. The resulting data and functional maps of the human genome provide a framework to assess the potential for non-coding mutations to dysregulate genes, and ultimately cellular regulation.

We first integrated diverse assays from ENCODE to define high-confidence regulatory elements and their associated genes, thereby defining what we call an extended gene neighborhood around each element. We also developed a regression based method for background mutation rate calibration, which removes confounding effects from chromatin and replication timing. Using this, we then search for genes with higher than expected mutational frequencies within the extended gene neighborhood. This approach was applied to successfully identify novel highly mutated genes that are associated with patient prognosis, such as such as *BCL6* in leukemia.

In addition, we also integrated extensive binding profiles from ENCODE to build tissue specific regulatory networks for both transcription factors (TFs) and RNA-binding proteins (RBPs). Intriguingly, when analyzing a hierarchical network, we found that TFs with higher mutation burden tend to be located at the bottom of the hierarchy (e.g., EZH2 and NR2C2), whereas those with dysregulated expression tend to reside at the top. Furthermore, by comparing the tumor-derived network with that from normal cells, we identified highly "rewired" TFs with changed targets and prognostic value, such as IKZF1 and MYC. We then extended the tissue specific network to build generalized networks across cancers. After combining this with expression profiles from other cohorts, we pinpointed MYC and SUB1 as key regulators that significantly drive tumor vs normal differential expression. We then validated their effects through knockdown experiments.

Finally, we propose a prioritization scheme for identifying key mutations in cancer. We identified active enhancers (and seven high impact-mutations therein) in breast cancer, and then validated their functional effects through luciferase assays.

Introduction

Mutations associated with cancer have been well characterized in key oncogenes and tumor suppressors. However, the overwhelming bulk of mutations in cancer genomes – particularly those discovered over the course of recent large-scale cancer genomics initiatives – lie within non-coding regions. Whether these mutations drive cancer development or progression, or simply emerge as byproducts of genomic instability remains an open question.

[[JZ2MG: I am not sure why the "However" sentence is deleted. Usually we should say someone has done some job but there is a lot of holes, so we did something else]]

Several recent studies have begun to address this question by employing a small repertoire of non-coding annotations and incorporating limited functional genomics data for variant interpretation \{cite 25261935, 27064257, 27807102}. For example, Weinhold *et al.* investigated recurrent non-coding mutations in regulatory regions. Wright *et al.* found cancer risk-associated single-nucleotide variants (SNVs) in enhancer regions that potentially upregulate MYC expression in colorectal cancer \{cite 20065031}. Lawrence *et al.* incorporated large scale genomics profiles to identify cancer drivers \{cite 23770567}.

[[JZ2MG: I know that to use "cancer-relevant ENCODE encyclopedia companion resource" is good in some sense, but they are way too long. I prefer the shorter name C-ENCODE]]

One way to approach the functional interpretation noncoding variants is to experimentally evaluate the functional effects mutating individual bases. This is a major endeavor of the ENCODE Consortium. In the initial release of the ENCODE annotation set five years ago, this was predominantly accomplished using RNA-Seq and ChIP-Seq assays on a limited number of cell lines. The new release of ENCODE took two new directions. First, it considerably broadened the number

Style Definition: No Spacing,TextBody: Font:9 pt

Comment [Office1]: This title is ok as is, but I feel like it is not that imaginative or catchy.

Deleted: mutations in the DNA…enomic variants which that disrupt the normal physiology of living cells. While Though mutations on …n coding genes have been well characterized, the preponderance of mutations in tumors tumor genomes occur …all within in …on-coding regions, rendering them less amenable to characterization and interpretation and are still poorly understood… The n…ewly released of ... [1]

Comment [j2]: Debating whether to put this here

Deleted: In this paper, w…e first integrated diverse assays from ENCODE to define high-confidence regulatory elements and their associated genes, thereby defining linkage to define the …hat we call an extended gene ... [2]

Deleted: , such as *BCL6* in leukemia, that are associ ... [3]

Comment [Office3]: What exactly doe you me ... [4]

Deleted: …pecific network to build up ... [5]

Comment [Office4]: True? (ie, is my change c ... [6]

Deleted: with …xpression profiles from other coho ... [7]

Deleted: d…a prioritization scheme for identifying ... [8]

Deleted: from the

Formatted: Highlight

Deleted: begin

Deleted: either directly

Deleted: small …mall repertoire group ... [9]

Deleted: or …nd incorporating limited functional ... [10]

Deleted: mutation effect

Formatted: Font:Italic

Deleted: like promoters and discovered mutations ... [11]

Formatted: Font:Italic

Deleted: variation …ariants (SNVs) in enhancer r ... [12]

Deleted: through long range interactions

Formatted: Font:Italic

Deleted: several expression, chromatin, and replic ... [13]

Deleted:

Deleted: quantify somatic mutation burden and

Comment [Office5]: I know Mark suggested ... [14]

Formatted: (Asian) Chinese (PRC), Highlight

Deleted: of …oncoding variants functional interpr ... [15]

Comment [Office6]: Insert citation here?

Formatted: Not Highlight

Deleted: went into

Formatted ... [16]

Deleted: On one hand

of cell lines with RNA-Seq and ChIP-Seq assays; the general ENCODE encyclopedia aims to utilize this to provide unified annotations. Secondly, ENCODE expanded the number of sophisticated assays such as STARR-seq, Hi-C, ChIA-pet, eCLIP and RAMPAGE on several top tier cell lines, many of which are cancer associated. This enables precise definitions of enhancers, direct identification of enhancement targets, and the construction of RNA binding protein (RBP) networks. Here, we focus on top-tier cell lines and their relevance to cancer by performing large-scale integration of these various assays to construct a companion resource to the general encyclopedia. We call this the "cancer-relevant ENCODE encyclopedia companion resource" (or the "encyclopedia companion" for short) for interpreting the wealth of mutational and transcriptional profiles produced by the cancer research community.

Data for comprehensive functional characterization in ENCODE

The most comprehensive set of assays for ENCODE are available for top tier cell lines. They provide good models not only for studying gene regulation in detail, but also for understanding cancers of the blood (K562), breast (MCF-7), liver (HepG2), lung (A549), and cervix (HeLa-S3). For four of these five top tier cell lines, there is another immortalized cell line from corresponding healthy tissue. Therefore, comparisons of the data between cancer and normal cell lines could help to model the differential gene regulation between tumor and normal tissues. It is worth noting that both relating these cell lines to cancers and pairing the tumor-normal matches are very approximate in nature, as these matches are not intended to substitute data from real tumor and normal tissues. Nonetheless, they serve as good models for performing a wide variety of functional genomics profiles, perturbation assays, and experimental validations. Furthermore, many of these pairings have been used in previous analyses \{cite 26241649 25144821}

(Fig 1A).

To build the encyclopedia companion, we started by defining enhancers. Unlike the general encyclopedia, we used a battery of 5 to 10 histone modification marks and DNase-seq signals. These were used as input into Casper, a machine learning predictor that we developed. Casper integrates the signal shapes of these various ChIP-seq and DNase-seq signals around enhancers. We then ensembled these predictions with peaks called from STARR-Seq experiments, which directly reads out candidate enhancers in the genome. Such an integrative approach gives accurate definitions of enhancers (see supplement). We then used RAMPAGE data to better define promoters, and further linked enhancers to putative promoters using a prediction algorithm based on gene expression. These potential linkages were then filtered through the results of Hi-C and ChIA-pet experiments, which provide a more accurate map of physical interactions to obtain high confidence enhancer target linkages. The enhancer target linkages, promoters, and RNA binding sites within genes constitute a so-called extended gene neighborhood (Fig1 C).

We further linked the enhancers and promoters with their associated transcription factors (TF) to construct extended regulatory networks. First, we built tissue specific distal and proximal TF regulatory networks by linking TF to genes either by direct TF-promoter interactions or indirection TF-enhancer-gene interactions (Fig1 B). We then pruned these networks to include only the strongest edges using another signal shape algorithm called TIP \{cite 22039215}. We investigated the rewiring status of these TFs in paired tumor-normal cell lines and made it available for download. In addition, we merged the tissue specific network to get a generalized network for pan-cancer analysis. We arranged all regulators in a hierarchy by a simulated procedure to give TFs with more than expected outbound edges at the top of hierarchy. TFs in different levels of the hierarchy reflect the extent to which they directly regulate the expression of other TFs \{cite 25880651}. A final hierarchal network structure is shown in Fig1 D.

Multi-level data integration better enables recurrent variant analysis in cancer

One of the most powerful ways of identifying key elements and functional mutations in cancer is with recurrence analysis to discover regions with greater mutational burden than expected. However, mutational processes can be influenced by confounding factors (in the form of both external genomic factors and local context effects), which can result in many false positives or negatives without appropriate correction \{cite MutsigCV}. In addition, traditional methods often neglect the natural association of mutations among different types of annotation and evaluate regions separately. Consequently, they sometimes fail to identify mutational signals from distributed yet biologically relevant genomic regions, thereby limiting interpretation.

To address these limitations, we adopt a two-pronged approach for integrating the encyclopedia companion for better recurrence analysis. First, we predict an accurate local background mutation rate (BMR) by removing effects of confounding factors in a cancer-specific manner. Specifically, we separated the whole genome into 1mb bins and calculated mutation counts under each local context category. For each category, we used a negative binomial regression of the mutation counts against features like replication timing, chromatin accessibility, Hi-C signals, and expression profiles for BMR prediction. In contrast to methods that use unmatched data \{cite MutsigCV}, our approach automatically selects the most relevant features, thereby providing noticeable improvements in BMR estimation, which significantly benefits recurrence analyses (Fig 2A).

Formatted ... [17]
Deleted: On the other hand…econdly, ENCODE a ... [18]
Formatted ... [19]
Deleted: …tier cell lines, many of which are canc ... [20]
Formatted ... [21]
Deleted: However there is no systematical integra ... [22]
Formatted ... [23]
Deleted: …ier cell lines. They provide good mod ... [24]
Comment [Office7]: Might just want to be ca ... [25]
Deleted: in …etween tumor …umorversus …and ... [26]
Deleted: [[REFs]].
Deleted: [JZ2MG: logic: raw data -> gene level -> ... [27]
Deleted: resource
Formatted ... [30]
Deleted: first
Formatted ... [31]
Deleted:
Formatted ... [32]
Deleted: like
Formatted ... [28]
Formatted ... [29]
Deleted: is with
Formatted ... [33]
Deleted: We integrated t
Formatted ... [34]
Formatted ... [35]
Deleted: called
Formatted ... [36]
Deleted: which
Formatted ... [37]
Comment [Office8]: Is there a better term for ... [40]
Deleted: data
Formatted ... [38]
Deleted:
Formatted ... [39]
Formatted ... [41]
Deleted:
Formatted ... [42]
Deleted: The w
Formatted ... [43]
Deleted: punitive
Formatted ... [44]
Deleted: prediction
Formatted ... [45]
Deleted:
Formatted ... [46]
Deleted:
Formatted ... [47]
Deleted: get
Formatted ... [48]
Deleted: …The enhancer- ... [49]
Formatted ... [50]
Deleted: n
Formatted ... [51]
Formatted ... [52]
Formatted ... [53]
... [54]
Formatted ... [55]
... [57]
Formatted ... [56]
Formatted ... [58]
Formatted ... [59]
... [60]

Rather than separately testing standalone annotation categories, we tested our extended gene neighborhoods as joint test units that contain both the coding exons and non-coding regulatory elements (Fig 1C). Such a scheme allows for the accumulation of weak mutational signatures distributed across multiple biologically relevant functional elements, which may otherwise be missed if evaluated under individual tests. We demonstrate that our scheme can effectively remove false positives and discover meaningful regions with higher-than-expected mutation counts (Fig 2C). For example, in the context of chronic lymphocytic leukemia (CLL), our analysis identifies well-known highly mutated genes, such as TP53 and ATM, which has been reported from previous coding region analysis. It also discovered genes that are missed by the exclusive analysis of coding regions, such as BCL6. Note that BCL6 has strong prognostic value with respect to patient survival (Fig. 2D), indicating that the extended gene neighborhood could be used as an annotation set for recurrence analysis. In addition, we can easily generalize this BMR calibration approach for other cancer types beyond the five discussed here, as our model will work to pick an appropriately matched ENCODE signal type.

Extensive rewiring events in cell-type specific network in cancer

We then investigated the transcriptional regulatory network in a cell-type specific way. Here, we utilized the 4 paired cancer-normal cell lines to study how the targets of each common TF changed (i.e., rewired) over the course of oncogenic transformation. We first ranked TFs according a "rewiring index" (Fig. 3 A), which calculates their respective number of lost and gained edges. In leukemia, oncogenes such as MYC and NRF1 are among the top edge gainers, while IKZF1, somatic mutations in which serve as a hallmark of high-risk acute lymphoblastic leukemia, is the most significant edge loser (Fig 3A). In contrast, several ubiquitously distributed TFs, such as YY1, retain their regulatory linkages (Fig 3A). We observe a similar trend in TFs using a distal, proximal and combined network (see details in supplementary file). We also observe highly rewired TFs such as BHLHE40, JUND, and MYC in lung, liver, and breast cancers (Fig 3).

Our rewiring index only considers direct connections associated with a given TF. However, the targets within the TF regulatory network are characterized by heterogeneous network modules (so called "gene communities"), which are dynamic and usually come from multiple biologically relevant genes. Instead of directly measuring the TF's target changes for each gene, we grouped these gene communities together to built a mixed-membership model. This enabled us to evaluate each TF's overall association changes to these gene communities in tumor and normal cells. Similar patterns are observed using this model (Fig 3A). To study the consequences of network rewiring under this model, we performed survival analysis on AML patients and found IKZF1 to be significantly associated with prognosis.

A remaining uncertainty lies in the identities of the factors associated with such massive rewiring events. We find that the majority of rewiring events are associated with noticeable gene expression and chromatin status changes, but not necessarily with variant-induced motif loss or gain events (Fig. 3A). For example, JUND is a top gainer in CLL. The majority of its gained targets in tumor cell lines demonstrate higher gene expression, stronger active and weaker repressive histone modification mark signals. We found a similar trend for the rewiring events associated with JUND in liver cancer.

Integrating regulatory networks with tumor expression profiles identifies key regulators in cancer

Next, we extended our network analysis in a pan-cancer way by merging these cell type specific networks for both TFs and RBPs. Using a machine learning method, we integrated 8,202 tumor expression profiles from TCGA to systematically search for TFs and RBPs that drive tumor-specific expression patterns. For each patient, our method tests whether a regulators' tumor to normal expression changes are sufficiently correlated with their targets' expression changes. We then calculated the percentage of patients with these relationships in each cancer type and presented the overall trends for key TFs and RBPs in Fig. 4A.

We find that the target genes of MYC are significantly up-regulated in numerous cancers, which is consistent with the known role of MYC as an oncogenic TF \cite{22464321}. We further validate MYC's regulatory effect through CRISPi RNA-seq experiments (Fig 4). Consistent with our predictions, the expression of MYC targets is significantly reduced after MYC knockdown (Fig 4A). After confirming the importance of MYC, we use the regulatory network to understand how MYC works with other TFs to accomplish this regulation. We first looked at all 3,735 triplets involving MYC by requiring that a second TF both interacts and shares a common target with MYC. In all cancer types, we found that MYC expression is positively correlated with the expression of most targets, while the second TF shows limited influence and random pattern as determined from partial correlation analysis. We then investigated the exact structure of such regulatory relationships in greater detail. The most common triplet interaction type is a well-understood feed-forward loop (FFL) in which MYC regulates both the common target and the second TF. Most of these MYC-associated FFLs are involved with well-known MYC partners such as Max and Mxl1. However, we also discovered that many involve another factor called NRF1, which has not been previously associated with MYC. Upon further study, we found that that the MYC-NRF1 FFL relationships are mostly coherent FFLs in which NRF1 amplifies the effect of MYC. We further studied the MYC-NRF1 FFLs by constituting the triplets into a logical gate, in which the two TFs act as inputs and the target gene expression represents the output \{cite 25884877}. We can show that the predominant number of

Deleted: which…contains…ontain both the codin ... [84]
Comment [Office13]: What do you mean by ... [85]
Deleted: new …enes that are missed by the exclus ... [86]
Comment [Office14]: Change to four?
Deleted: [[integ better]]
Formatted ... [87]
Deleted: not in
Comment [Office15]: Sounds a bit awkward ... [88]
Deleted: we are focusing on
Deleted: . ... [89]
Comment [Office16]: Change to one of the ... [90]
Deleted: regulation
Deleted: tissue
Deleted: In each cell type, …ere, we utilized the 4 ... [91]
Moved up [1]: we organized the TF regulatory ne ... [93]
Formatted ... [94]
Moved up [2]: TFs in different levels of the hiera ... [95]
Deleted: [[cut]]we organized the TF regulatory ne ... [92]
Formatted ... [96]
Deleted: Specifically, for leukemia we removed th ... [97]
Deleted: and …iver, and breast cancers (see fig X ... [98]
Deleted:) although we do not have as many comn ... [99]
Deleted: [[In particular,…]]
Deleted: Our rewiring index only considers direc ... [100]
Formatted ... [101]
Formatted ... [102]
Formatted ... [103]
Deleted: in
Formatted ... [104]
Deleted: is
Formatted ... [105]
Deleted: the
Formatted ... [106]
Deleted: , the
Formatted ... [107]
Deleted: y
Formatted ... [108]
Deleted: ,…, which is ... [109]
Formatted ... [110]
Comment [Office19]: What is meant by this?
Formatted ... [111]
Deleted: …nstead of directly measuring the TF' ... [112]
Comment [Office20]: This language does r ... [113]
Formatted ... [114]
Deleted: .
Comment [Office21]: Change to "with poor ... [115]
Formatted ... [116]
Deleted: .
Deleted: associated …actors associated of ... [117]
Deleted: changes in
Deleted: K562
Deleted: and …he majority of its gained or lost ... [118]
Deleted: s… stronger active but ... [119]
... [120]
... [121]
Formatted ... [122]
Formatted ... [123]
... [125]
Formatted ... [124]
Formatted ... [126]
Formatted ... [127]

these logic gates follow either an OR gate or a gate in which MYC is always dominant. This describes the regulatory logic implicit in the cancers in great detail. It also illustrates the power of the regulatory network to not only find key regulators but also to really demonstrate how they work in combination with other regulators.

We also analyzed the RBP network derived from ENCODE data and found key RBPs associated with cancer. For example, the ENCODE eCLIP experiment has profiled many SUB1 peaks on the 3'UTR regions of genes, and we find that the predicted targets of the RBP SUB1 were significantly up-regulated in many cancer types (Fig. 4C). Specifically, in liver cancer, after knocking down SUB1 in HepG2 cells, its predicted targets are also down-regulated relative to other genes (Fig. 4D). In addition, we found that the decay rate of SUB1 target genes are significantly shorter than non-targets (Fig. 4C). These results indicate that SUB1 may bind to 3'UTR regions to stabilize transcripts. From our integrated analysis, the higher SUB1 activity through regulatory binding on 3'UTR regions is likely to drive tumor-specific expression patterns in many cancer types. Moreover, the up-regulation of SUB1 target genes is correlated with a poorer patient survival in cancer types such as lung cancer (Fig. 4).

Step-wise prioritization schemes pinpoint deleterious SNVs in cancer

In summary, our cancer encyclopedia constitutes a list of regions with higher than expected mutational frequencies in cancer, accurately determined enhancers and gene linkages, extended gene neighborhoods, a regulatory network of TFs (and for some lines, a regulatory RBP network). The cancer encyclopedia also provides the characteristics of TF/RBPs within the network, such as positions within the network hierarchy, rewiring status, and tumor/normal differential expression in driving cancer, along with prognostic values in various cancer types. Collectively, these resources allow us to prioritize a few key elements as being associated with oncogenesis.

The workflow in Fig. 5A describes this prioritization scheme in a systematic fashion. We first search for key regulators that are frequently rewired, located in network hubs or at top of the network hierarchy, or significantly drive expression changes in cancer. We then prioritize functional elements that are associated with top regulators, undergo large regulatory changes in terms of TF binding and chromatin status, or are highly mutated in tumors. Finally, on a nucleotide level, we can pinpoint impactful SNVs for small-scale functional characterization by their ability to disrupt or create specific binding sites, or which occur in positions of particularly high conservation.

[[JZ2MG: the first sentence is very confusing, seem that these enhancers are associated with these TFs. I suggest we just remove the 1st sentence. Also we mentioned this in detail already]]

Using this framework, we subject a number of key regulators, such as MYC and SUB1, to knockdown experiments to validate their regulatory effects (Fig 4D). We also identified several active enhancers in noncoding regions, and validated their ability to influence transcription using luciferase assays. We further selected key SNVs within these enhancers that are important for controlling gene expression. Of the eight motif-disrupting SNVs that we tested, six showed consistent up- or down-regulation relative to the wild type. One particularly interesting example, illustrating the unique value of ENCODE data integration, is in the intronic region of CDH26 in chromosome 20 (Fig. 5C). Both histone modification and chromatin accessibility (DNase-seq) signals indicated an active regulatory role, which was further confirmed as an enhancer by both CASPER predictions and ESCAPE peaks on STARR-seq (Fig. 5D). Hi-C and ChIA-PET data indicated that the region is within a topologically associated domain (TAD) and showed a regulatory linkage to the downstream cancer associated gene SYCP2 \cite{26334652, 24662924}. We found that 21 out of the 52 ChIP-Seq experiments in MCF-7 from ENCODE demonstrate a high frequency of chromatin interactions in this region. Motif analysis predicts that C-to-G mutations in cancer can significantly disrupt the binding affinity of several TFs, such as FOLS2. Luciferase assays demonstrate that this mutation introduces an xx-fold reduction in expression relative to wild type expression levels, indicating a strong repressive effect on this enhancer's functionality.

Conclusion

This study highlights the value of our encyclopedia companion as a resource for cancer research. Our encyclopedia companion is leveraged to provide a prioritization scheme for pinpointing key regulatory elements and SNVs for small-scale validations. One of the key aspect of this study is that, by integrating many different types of data on a large scale, we demonstrate how we may achieve an accurate annotation of cell lines that draws into cancer. By integrating this with expression and mutation data from patient samples, we can highlight key regions and regulators involved in oncogenesis. This study underscores the value of large scale data integration, and we note that expanding this approach (either by integrating additional data types and/or using tumor mutation and expression data on a larger scale) is straightforward. We also anticipate that an additional step would be to determine a lot of the ENCODE assays on specific tissues and tumor samples. Though volume of material needed for such analyses may present challenges, we show that such a framework is technically feasible and provides further opportunities for the future.

Formatted ... [171]
Deleted: …This describes the regulatory logic i ... [172]
Deleted: not
Deleted: in …ind ing ... [175]
Formatted ... [173]
Formatted ... [174]
Formatted ... [176]
Deleted: really…o really demonstrate showing ... [177]
Formatted ... [178]
Deleted:
Deleted: .
Deleted: In addition to recapitulating existing kn ... [179]
Deleted: The TF analysis shows that we are succ ... [180]
Moved (insertion) [4] ... [181]
Deleted: …or example, the ENCODE eCLIP ex ... [182]
Moved (insertion) [5] ... [183]
Deleted: Previously, SUB1 was considered as a ... [184]
Deleted: gene…o 3'UTR regions to stabilize tra ... [185]
Moved (insertion) [6] ... [186]
Deleted: worse
Comment [Office26]: maube change "canc ... [187]
Deleted: …. . ... [188]
Moved (insertion) [7] ... [189]
Deleted: [[moved down]]
Formatted ... [190]
Formatted ... [191]
Moved (insertion) [3] ... [192]
Deleted: consists …onstitutes a list of regions w ... [193]
Comment [Office27]: You might want to do ... [194]
Moved up [4]: For example, the predicted targ ... [196]
Deleted:
Moved up [5]: Previously, SUB1 was considere ... [197]
Deleted: .
Deleted: .
Moved up [7]: . ... [198]
Deleted: , some of which are then validated usin ... [195]
Deleted: .
Deleted: in
Deleted: changes… or (most importantly) ... [199]
Formatted ... [200]
Formatted ... [201]
Deleted: then
Deleted: control… Of the eight motif-disrupting ... [202]
Deleted: on …n the intronic region of CDH26 in ... [203]
Deleted:
Comment [Office29]: Is this written like "CA ... [204]
Deleted: …i-C and ChIA-PET data indicated tha ... [205]
Deleted: This enhancer is located in a noncoding ... [206]
Deleted: the C …-to- … mutations in cancer can ... [207]
Deleted: FOLS2
Comment [Office30]: ?
Formatted ... [208]
Deleted: values…alue of our cancer …ncyclope ... [209]
... [210]
... [211]
... [212]
... [213]
... [214]
Formatted ... [215]
Formatted ... [216]

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [1] Deleted **Microsoft Office User** **3/10/17 1:31:00 PM**

mutations in the DNA

Page 1: [2] Deleted **Microsoft Office User** **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted **Microsoft Office User** **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted **Microsoft Office User** **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted **Microsoft Office User** **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted **Microsoft Office User** **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted | **Microsoft Office User** | **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted | **Microsoft Office User** | **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted | **Microsoft Office User** | **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [2] Deleted | **Microsoft Office User** | **3/10/17 1:37:00 PM**

In this paper, w

Page 1: [3] Deleted | **Microsoft Office User** | **3/10/17 1:58:00 PM**

, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides,

Page 1: [3] Deleted | **Microsoft Office User** | **3/10/17 1:58:00 PM**

, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides,

Page 1: [3] Deleted | **Microsoft Office User** | **3/10/17 1:58:00 PM**

, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides,

Page 1: [3] Deleted | **Microsoft Office User** | **3/10/17 1:58:00 PM**

, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides,

Page 1: [3] Deleted | **Microsoft Office User** | **3/10/17 1:58:00 PM**

, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides,

Page 1: [4] Commented | **Microsoft Office User** | **3/10/17 2:05:00 PM**

What exactly doe you mean by this? I suggest revising.

Page 1: [5] Deleted | **Microsoft Office User** | **3/10/17 2:05:00 PM**

Page 1: [5] Deleted | **Microsoft Office User** | **3/10/17 2:05:00 PM**

Page 1: [6] Commented | **Microsoft Office User** | **3/10/17 2:06:00 PM**

True? (ie, is my change correct here?)

Page 1: [7] Deleted	Microsoft Office User	3/10/17 2:06:00 PM

with

Page 1: [7] Deleted	Microsoft Office User	3/10/17 2:06:00 PM

with

Page 1: [7] Deleted	Microsoft Office User	3/10/17 2:06:00 PM

with

Page 1: [7] Deleted	Microsoft Office User	3/10/17 2:06:00 PM

with

Page 1: [7] Deleted	Microsoft Office User	3/10/17 2:06:00 PM

with

Page 1: [8] Deleted	Microsoft Office User	3/10/17 2:09:00 PM

d

Page 1: [8] Deleted	Microsoft Office User	3/10/17 2:09:00 PM

d

Page 1: [8] Deleted	Microsoft Office User	3/10/17 2:09:00 PM

d

Page 1: [9] Deleted	Microsoft Office User	3/10/17 2:13:00 PM

small

Page 1: [9] Deleted	Microsoft Office User	3/10/17 2:13:00 PM

small

Page 1: [10] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 11:22:00 AM

or

Page 1: [10] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 11:22:00 AM

or

Page 1: [11] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 11:20:00 AM

like promoters and discovered mutations in promoter that reduce gene expression and suggests poor prognosis

Page 1: [12] Deleted	Microsoft Office User	3/10/17 2:16:00 PM

variation

Page 1: [12] Deleted	Microsoft Office User	3/10/17 2:16:00 PM

variation

Page 1: [13] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 11:21:00 AM

several expression, chromatin, and replication timing

Page 1: [14] Commented Microsoft Office User 3/10/17 2:18:00 PM

I know Mark suggested making this shorter, but I actually feel that it is too short. My opinion is that, if you do make it this short, it does more harm than good to mention these other papers in such brief terms. Thus, it may be better to simply list the citations for these papers, without the very brief synopsis on each of the papers (ie, the brief synopsis doesn't really do enough justice to these papers I feel)

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [15] Deleted Microsoft Office User 3/10/17 2:21:00 PM

of

Page 1: [16] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 11:59:00 AM

Not Highlight

Page 1: [16] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 11:59:00 AM

Not Highlight

Page 2: [17] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:02:00 PM

Not Highlight

Page 2: [18] Deleted Microsoft Office User 3/10/17 2:24:00 PM

On the other hand

Page 2: [18] Deleted Microsoft Office User 3/10/17 2:24:00 PM

On the other hand

Page 2: [19] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:00:00 PM

Not Highlight

Page 2: [19] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:00:00 PM

Not Highlight

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [20] Deleted Microsoft Office User 3/10/17 2:25:00 PM

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:12:00 PM

Not Highlight

Page 2: [21] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:12:00 PM

Not Highlight

Page 2: [22] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 11:20:00 AM

However there is no systematical integration of thousands of functional genomic data sets from tens of experimental assays to interpret the cancer genome.[[shorten but keep refs]]

The newly-released data from the ENCODE Consortium can benefit such integrative analysis by providing comprehensive characterization of non-coding regulatory elements and linking them to cancer associated genes. The second phase of ENCODE was focusing on using RNA-seq and ChIP-seq data from multiple cell lines to define non-coding regulatory elements. Phase three ENCODE went into two directions. On one hand, it expanded the cell lines and tissue for these RNA-seq and ChIP-seq data to get a general catalog of regulatory element, which has been covered in the main ENCODE encyclopedia paper; on the other hand, focusing on the top tier cell line it expanded the number of sophisticated assays such as STARR-seq, Hi-C, ChIA-pet, and RAMPAGE. These improvements enable us to accurately identify distal regulatory elements such as enhancers and link them to genes, whereas the broader catalogue gives more general set of regulatory elements in many tissues. Here, we endeavor to provide a companion resource to the main ENCODE encyclopedia by focusing on cancer and building aan "cancer-relevant ENCODE encyclopedia". companion resource" (or encyclopedia companion for short). The main encyclopedia is oriented toward breath of the annotations to describe elements over hundreds of cell lines. In contrast, we focus on top tier cell lines with a wide variety of assays available. Most of these cell lines are associated with cancers of the blood, liver, lung, cervix, and breast. We show that these cell lines can be used to provide a better understanding of oncogenesis, and we provide a resource for interpreting the wealth of mutational and transcriptional profiles produced by the cancer community.

Page 2: [23] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 2: [23] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 2: [23] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 2: [23] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 2: [23] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 2: [24] Deleted	Microsoft Office User	3/10/17 2:27:00 PM

Page 2: [24] Deleted	Microsoft Office User	3/10/17 2:27:00 PM

Page 2: [24] Deleted	Microsoft Office User	3/10/17 2:27:00 PM

Page 2: [24] Deleted	Microsoft Office User	3/10/17 2:27:00 PM

Page 2: [25] Commented	Microsoft Office User	3/10/17 2:35:00 PM

Might just want to be careful w/the language. I understand that it is “normal” in the sense that it is not a cancer cell line, but by being immortalized, a reviewer might object to the term “normal” here.

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [26] Deleted	Microsoft Office User	3/10/17 2:36:00 PM

in

Page 2: [27] Deleted	jingzhang.wti.bupt@gmail.com	3/10/17 10:54:00 AM

[JZ2MG: logic: raw data -> gene level -> network level]

Page 2: [28] Formatted	jingzhang.wti.bupt@gmail.com	3/10/17 10:19:00 AM

No Spacing,TextBody, Left, Line spacing: single

Page 2: [29] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [30] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [31] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [32] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [33] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [33] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [33] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [33] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [34] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [34] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [35] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [36] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [37] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [37] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [37] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [37] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [38] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [39] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [39] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [40] Commented	Microsoft Office User	3/10/17 3:04:00 PM

Is there a better term for this? The term “ensemble” is a noun, and I don’t think it is a verb (ie, I don’t think one can use the term “ensembled”, but I could be mistaken)

Page 2: [41] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [41] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [41] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [42] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [42] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted	jingzhang.wti.bupt@gmail.com	3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [43] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [44] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [44] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [44] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [45] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [46] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:12:00 PM

Not Highlight

Page 2: [46] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:12:00 PM

Not Highlight

Page 2: [47] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:12:00 PM

Not Highlight

Page 2: [47] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:12:00 PM

Not Highlight

Page 2: [48] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [48] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM

Font:Times New Roman

Page 2: [49] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [49] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [50] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [50] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [50] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [50] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [50] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [51] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [51] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 12:59:00 PM
Font:Times New Roman
Page 2: [52] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [52] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [52] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [53] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [54] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [54] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [54] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [55] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [55] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [55] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM
Font:Times New Roman
Page 2: [56] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [56] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [56] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [57] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [57] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [58] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [59] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [60] Deleted Microsoft Office User 3/10/17 2:16:00 PM

Page 2: [60] Deleted Microsoft Office User 3/10/17 2:16:00 PM

Page 2: [61] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [61] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [62] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [62] Deleted Microsoft Office User 3/10/17 1:54:00 PM

Page 2: [63] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [63] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:23:00 PM

Font:Times New Roman

Page 2: [64] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:36:00 PM

To build a cancer encyclopedia[j1] with these cell lines, we first collected comprehensive [j2] functional genomics data to characterize factors that potentially affect somatic mutagenic processes. Then at the gene level, we tried to accurately identify both distal (enhancers) and proximal (promoters and regulator binding sites) regulatory elements. Specifically, we first focused on identifying enhancers and linking them to genes through an ensemble method (Fig. S2). In contrast to methods relying on a single assay, we first used a pattern recognition based algorithm called CASPER on ChIP-seq and DNase-seq signals to search for enhancer candidates and then pruned them using peaks from our STARR-seq pipeline ESCAPE. We further applied our enhancer linkage prediction method JEME based on ChIP-seq, DNase-seq, ChIA-pet, and RNA-seq data to link these enhancers to genes.

These potential linkages were then filtered through the results of Hi-C experiments, which provide a more accurate yet lower resolution map of chromatin interactions. For each gene, we combined these enhancers with proximal regulatory elements to construct what we termed “extended gene neighborhoods” – coding regions matched with key regulatory elements – to better interpret gene regulation (Fig1 B). Furthermore, at the network level we also explored the binding profiles in ENCODE and constructed high-confidence gene regulatory networks for both TFs and RBPs (Fig1 C). Finally, we merged our efforts with the broader ENCODE encyclopedia and provided consistent identifiers and definitions for the cancer encyclopedia. [[REWIRING]]

we organized the TF regulatory network into a hierarchy by comparing the inbound and outbound edges of each factor, thereby enabling us to investigate the global topology of TF regulation (Fig. 1E, see also Supp. File/Section(?) X).

TFs in different levels of the hierarchy reflect the extent to which they directly regulate the expression of other TFs \{cite 25880651}. For example, TFs in the top layer have more outbound than inbound edges in the network, and thus play larger roles in regulating other TFs (Supp. Fig. xx). In this representation, two patterns readily emerge. In leukemia, top-level TFs tend to more strongly influence the differential expression between tumor and normal cells. The average Pearson correlation between TF binding events and tumor-normal expression changes increases from 0.125 in the bottom layer to 0.270 in the top layer (Table Sx). TFs in the bottom layer are more frequently associated with burdened [SL3]binding sites in general, perhaps reflecting their increased resilience to mutation [SL4](see Supp. Section X, Table Sx).

Page 2: [65] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Highlight

Page 2: [66] Moved from page 3 (Move #1) **Microsoft Office User** **3/9/17 11:17:00 AM**

we organized the TF regulatory network into a hierarchy by comparing the inbound and outbound edges of each factor, thereby enabling us to investigate the global topology of TF regulation (Fig. 1E, see also Supp. File/Section(?) X).

Page 2: [67] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Font color: Auto, Highlight

Page 2: [68] Moved from page 3 (Move #2) **Microsoft Office User** **3/9/17 11:17:00 AM**

TFs in different levels of the hierarchy reflect the extent to which they directly regulate the expression of other TFs \{cite 25880651}. For example, TFs in the top layer have more outbound than inbound edges in the network, and thus play larger roles in regulating other TFs (Supp. Fig. xx). In this representation, two patterns readily emerge. In leukemia, top-level TFs tend to more strongly influence the differential expression between tumor and normal cells. The average Pearson correlation between TF binding events and tumor-normal expression changes increases from 0.125 in the bottom layer to 0.270 in the top layer (Table Sx). TFs in the bottom layer are more frequently associated with burdened [SL5]binding sites in general, perhaps reflecting their increased resilience to mutation [SL6](see Supp. Section X, Table Sx).

Page 2: [69] Moved to page 3 (Move #3) **Microsoft Office User** **3/9/17 11:17:00 AM**

In summary, our cancer encyclopedia consists a list of regions with higher than expected mutation frequency in cancer, accurately determined enhancers and gene linkages, the extended gene neighborhoods, the regulatory network of TFs (and for some lines RBPs), as well as the characteristics of TF/RBPs within the network, such as positions in network hierarchy, rewiring status, tumor/normal differential expression driving potential and prognostic value in various cancer types. Collectively, these resources allow us to prioritize a few key elements as being associated with oncogenesis, some of which are then validated using small-scale experiments (see table S1).

Page 2: [70] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Highlight

Page 2: [71] Deleted **Microsoft Office User** **3/9/17 11:17:00 AM**

throughvia

Page 2: [71] Deleted **Microsoft Office User** **3/9/17 11:17:00 AM**

throughvia

Page 2: [72] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:50:00 PM

,

Page 2: [72] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:50:00 PM

,

Page 2: [72] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:50:00 PM

,

Page 2: [72] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:50:00 PM

,

Page 2: [73] Deleted Microsoft Office User 3/10/17 3:46:00 PM

that

Page 2: [73] Deleted Microsoft Office User 3/10/17 3:46:00 PM

that

Page 2: [74] Deleted Microsoft Office User 3/9/17 11:17:00 AM

could

Page 2: [74] Deleted Microsoft Office User 3/9/17 11:17:00 AM

could

Page 2: [75] Deleted Microsoft Office User 3/9/17 11:17:00 AM

categories

Page 2: [75] Deleted Microsoft Office User 3/9/17 11:17:00 AM

categories

Page 2: [76] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:51:00 PM

dispersed

Page 2: [76] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:51:00 PM

dispersed

Page 2: [77] Deleted Microsoft Office User 3/10/17 3:48:00 PM

integrated

Page 2: [77] Deleted Microsoft Office User 3/10/17 3:48:00 PM

integrated

Page 2: [78] Formatted Microsoft Office User 3/9/17 11:17:00 AM

Highlight

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [79] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:39:00 PM

BMR

Page 2: [80] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:57:00 PM

[[too detailed]]

Page 2: [80] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:57:00 PM

[[too detailed]]

Page 2: [81] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:59:00 PM

Not Highlight

Page 2: [81] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:59:00 PM

Not Highlight

Page 2: [81] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 2:59:00 PM

Not Highlight

Page 2: [82] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 2:59:00 PM

usually yields higher BMR prediction precision

Page 2: [82] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 2:59:00 PM

usually yields higher BMR prediction precision

Page 2: [83] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [84] Deleted	Microsoft Office User	3/10/17 3:54:00 PM

which

Page 3: [85] Commented	Microsoft Office User	3/10/17 3:55:00 PM

What do you mean by this? It seems a little awkward (and unclear) to me. Do you mean that this gene has been implicated in cancer in canonical/classical coding-based studies (ie, those mentioned in the 2nd sentence of the abstract)?

Page 3: [86] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

new

Page 3: [86] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

new

Page 3: [86] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

new

Page 3: [87] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:01:00 PM

Not Highlight

Page 3: [88] Commented Microsoft Office User 3/10/17 4:00:00 PM

Sounds a bit awkward and unclear

Page 3: [89] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:02:00 PM

Page 3: [90] Commented Microsoft Office User 3/10/17 4:01:00 PM

Change to one of the following:
"in cell-type specific cancer networks"
"in the cell-type specific cancer network"
"in the cell-type specific network in cancer"

Page 3: [91] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In each cell type,

Page 3: [91] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In each cell type,

Page 3: [91] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In each cell type,

Page 3: [91] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In each cell type,

Page 3: [91] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In each cell type,

Page 3: [92] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:03:00 PM

[[cut]]we organized the TF regulatory network into a hierarchy by comparing the inbound and outbound edges of each factor, thereby enabling us to investigate the global topology of TF regulation (Fig. 1E, see also Supp. File/Section(?) X). TFs in different levels of the hierarchy reflect the extent to which they directly regulate the expression of other TFs \{cite 25880651}. For example, TFs in the top layer have more outbound than inbound edges in the network, and thus play larger roles in regulating other TFs (Supp. Fig. xx). In this representation, two patterns readily emerge. In leukemia, top-level TFs tend to more strongly influence the differential expression between tumor and normal cells. The average Pearson correlation between TF binding events and tumor-normal expression changes increases from 0.125 in the bottom layer to 0.270 in the top layer (Table Sx). TFs in the bottom layer are more frequently associated with burdened [SL7]binding sites in general, perhaps reflecting their increased resilience to mutation [SL8](see Supp. Section X, Table Sx).

When comparing the common regulators in approximately matched tumor and normal regulatory networks, rewiring (i.e., target changing) analysis may help to identify cancer-associated deregulation. Hence, we investigated rewiring events in TF networks using multiple formulations (see Supp. File/Section(?) X).

Page 3: [93] Moved to page 2 (Move #1) Microsoft Office User 3/9/17 11:17:00 AM

we organized the TF regulatory network into a hierarchy by comparing the inbound and outbound edges of each factor, thereby enabling us to investigate the global topology of TF regulation (Fig. 1E, see also Supp. File/Section(?) X).

Page 3: [94] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Highlight

Page 3: [95] Moved to page 2 (Move #2)	Microsoft Office User	3/9/17 11:17:00 AM

TFs in different levels of the hierarchy reflect the extent to which they directly regulate the expression of other TFs \{cite 25880651}. For example, TFs in the top layer have more outbound than inbound edges in the network, and thus play larger roles in regulating other TFs (Supp. Fig. xx). In this representation, two patterns readily emerge. In leukemia, top-level TFs tend to more strongly influence the differential expression between tumor and normal cells. The average Pearson correlation between TF binding events and tumor-normal expression changes increases from 0.125 in the bottom layer to 0.270 in the top layer (Table Sx). TFs in the bottom layer are more frequently associated with burdened [SL9]binding sites in general, perhaps reflecting their increased resilience to mutation [SL10](see Supp. Section X, Table Sx).

Page 3: [96] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Font color: Red, Highlight

Page 3: [96] Formatted	Microsoft Office User	3/9/17 11:17:00 AM

Font color: Red, Highlight

Page 3: [97] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

Specifically, for leukemia we removed the general TFs and restricted our rewiring analysis to 61 common TFs in K562 and GM12878 from ENCODE.

Page 3: [97] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

Specifically, for leukemia we removed the general TFs and restricted our rewiring analysis to 61 common TFs in K562 and GM12878 from ENCODE.

Page 3: [97] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

Specifically, for leukemia we removed the general TFs and restricted our rewiring analysis to 61 common TFs in K562 and GM12878 from ENCODE.

Page 3: [97] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

Specifically, for leukemia we removed the general TFs and restricted our rewiring analysis to 61 common TFs in K562 and GM12878 from ENCODE.

Page 3: [97] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

Specifically, for leukemia we removed the general TFs and restricted our rewiring analysis to 61 common TFs in K562 and GM12878 from ENCODE.

Page 3: [98] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 3:07:00 PM

and

Page 3: [98] Deleted	jingzhang.wti.bupt@gmail.com	3/9/17 3:07:00 PM

and

Page 3: [99] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

) although we do not have as many common TFs between tumor and normal cell lines for these tissues.

Page 3: [100] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:52:00 AM

Our rewiring index only considers direct connections associated with a given TF. One may also consider rewiring that include not only direct connections, but also the whole neighborhood of connections with which a TF associates through membership and topic models. In particular, we used a mixed-membership model to look more abstractly at local gene neighborhoods to re-rank the TFs (see Supp. File/Section(?) X). Similar patterns are observed using this model. We also observed that MYC (a well-known oncogene) becomes a top edge gainer (Fig 3A). To study the consequences of network rewiring under this model, we performed survival analysis on xxx AML patients and found IKZF1 to be significantly associated with prognosis.

Page 3: [101] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:15:00 PM

Highlight

Page 3: [102] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:53:00 AM

No Spacing,TextBody

Page 3: [103] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [104] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [104] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [105] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [106] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [107] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [107] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [108] Formatted jingzhang.wti.bupt@gmail.com 3/10/17 10:42:00 AM

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [109] Deleted Microsoft Office User 3/10/17 4:11:00 PM

,

Page 3: [109] Deleted Microsoft Office User 3/10/17 4:11:00 PM

,

Page 3: [110] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:42:00 AM**

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [111] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:42:00 AM**

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [111] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:42:00 AM**

Font:(Default) Times New Roman, (Asian) +Theme Body Asian (ＭＳ 明朝), 9 pt, Font color: Auto, Pattern: Clear

Page 3: [112] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [112] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [112] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [112] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [113] Commented **Microsoft Office User** **3/10/17 4:14:00 PM**

This language does not seem right to me, but I'm not 100% what is intended.

Page 3: [114] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:52:00 AM**

Not Highlight

Page 3: [115] Commented **Microsoft Office User** **3/10/17 4:15:00 PM**

Change to "with poor prognosis"? The term "prognosis" by itself is ambiguous.

Page 3: [116] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:52:00 AM**

Not Highlight

Page 3: [116] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:52:00 AM**

Not Highlight

Page 3: [117] Deleted **Microsoft Office User** **3/10/17 4:17:00 PM**

associated

Page 3: [117] Deleted **Microsoft Office User** **3/10/17 4:17:00 PM**

associated

Page 3: [118] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:17:00 PM**

and

Page 3: [118] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:17:00 PM**

and

Page 3: [118] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:17:00 PM**

and

Page 3: [119] Deleted **Microsoft Office User** **3/10/17 4:17:00 PM**

s

Page 3: [119] Deleted **Microsoft Office User** **3/10/17 4:17:00 PM**

s

Page 3: [120] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:17:00 PM**

substantial expression and chromatin status changes (at least 2-fold).

Page 3: [121] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:16:00 PM**

It is interesting to further investigate the causal relationship between such changes and the rewiring events.[JZ2MG: since we are not able to make a conclusion here, I just said something very vague, but maybe it is too vague…]…][[Not sure what you're getting at – shirley's comment]]

Page 3: [122] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [123] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [123] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [124] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [125] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:24:00 PM**

networks and performed a generalized pan-cancer generalized network analysis across multiple cancer types

Page 3: [125] Deleted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:24:00 PM**

networks and performed a generalized pan-cancer generalized network analysis across multiple cancer types

Page 3: [126] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [127] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [128] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [128] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [129] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 3:31:00 PM**

Not Highlight

Page 3: [130] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [131] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:30:00 PM

the regulatory targets

Page 3: [131] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:30:00 PM

the regulatory targets

Page 3: [132] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [133] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:26:00 PM

across tumors

Page 3: [133] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:26:00 PM

across tumors

Page 3: [134] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [135] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [136] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [136] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 3:31:00 PM

Not Highlight

Page 3: [137] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:28:00 PM

regulated between each pair of cancer type and regulator. The

Page 3: [137] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:28:00 PM

regulated between each pair of cancer type and regulator. The

Page 3: [137] Deleted jingzhang.wti.bupt@gmail.com 3/9/17 3:28:00 PM

regulated between each pair of cancer type and regulator. The

Page 3: [138] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:12:00 AM

The predicted impacts of regulators on tumor gene expression are highly consistent with previous findings.

Page 3: [139] Formatted Microsoft Office User 3/10/17 4:25:00 PM

No Spacing,TextBody, Left, Line spacing: single

Page 3: [140] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:01:00 AM

In addition, we further investigated how MYC interacted with other TFs to jointly control their target gene expressionsexpression. We found that genes shared with MYC and other TFs only showed marginal partial correlation with MYC's co-regulatory TFs,

indicating a major role of MYC for gene expression control. We then the particular network motif feed-forward loops (FFLs) with MYC as the master regulator. Among the 38 other TFs that form FFLs with MYC, we selected NRF1 since it has the most common targets except two well-known MYC partner TFs. We then checked the logic gate usage among MYC and NRF1 and found that either MYC itself or the or gates dominates across multiple cancer types, confirming the major role of MYC in gene regulation. We also discovered that these MYC-NRF1 FFLs are mostly coherent ones, where NRF1 serves as an amplifier to the MYC effects on the target genes. This is consistent with the previous discovery that NRF1 is intervening to MYC's apoptotic function \{cite 12533512 }.

Page 3: [141] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Font color: Text 1, Highlight

Page 3: [141] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Font color: Text 1, Highlight

Page 3: [142] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [142] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [142] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [142] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [143] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [143] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [144] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [145] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [146] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [146] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [147] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [147] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [148] Deleted **Microsoft Office User** **3/10/17 4:23:00 PM**

,

Page 3: [148] Deleted **Microsoft Office User** **3/10/17 4:23:00 PM**

,

Page 3: [149] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [149] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [149] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [149] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [150] Commented Microsoft Office User 3/10/17 4:24:00 PM

Consider deleting. It sounds very awkward, and I think it is redundant with “shows limited influence”

Page 3: [151] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [152] Commented Microsoft Office User 3/10/17 4:25:00 PM

Consider a paragraph break here.

Page 3: [153] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [153] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [154] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [154] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [155] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [155] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [155] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [156] Deleted Microsoft Office User 3/10/17 4:25:00 PM

where

Page 3: [156] Deleted Microsoft Office User 3/10/17 4:25:00 PM

where

Page 3: [157] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [157] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 3: [158] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [158] Deleted **Microsoft Office User** **3/10/17 1:54:00 PM**

Page 3: [159] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [159] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [159] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [160] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [160] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [160] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [161] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 3: [161] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 3: [162] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [162] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [162] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [162] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [163] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [164] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Furthermore

Page 3: [164] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Furthermore

Page 3: [165] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [165] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [165] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [165] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [166] Commented **Microsoft Office User** **3/10/17 4:28:00 PM**

Maybe use another term here. Something like “representing as” or something?

Page 3: [167] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [168] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [168] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [169] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 3: [169] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 4: [170] Deleted **Microsoft Office User** **3/10/17 4:28:00 PM**

sign

Page 4: [170] Deleted **Microsoft Office User** **3/10/17 4:28:00 PM**

sign

Page 4: [171] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 4: [172] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [172] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [173] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 4: [174] Formatted **jingzhang.wti.bupt@gmail.com** **3/9/17 4:46:00 PM**

Font:Times New Roman, Font color: Text 1

Page 4: [175] Deleted **Microsoft Office User** **3/10/17 4:29:00 PM**

in

Page 4: [175] Deleted Microsoft Office User 3/10/17 4:29:00 PM

in

Page 4: [176] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 4: [176] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 4: [177] Deleted Microsoft Office User 3/10/17 4:29:00 PM

really

Page 4: [177] Deleted Microsoft Office User 3/10/17 4:29:00 PM

really

Page 4: [178] Formatted jingzhang.wti.bupt@gmail.com 3/9/17 4:46:00 PM

Font:Times New Roman, Font color: Text 1

Page 4: [179] Deleted Microsoft Office User 3/9/17 11:17:00 AM

In addition to recapitulating existing knowledge from previous studies, our analysis also predicts previously unidentified functions for regulators in cancer.

Page 4: [180] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:01:00 AM

The TF analysis shows that we are successfully able to use the ENCODE regulatory network to identify and analyze key TFs in cancer gene expression data.

Page 4: [180] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:01:00 AM

The TF analysis shows that we are successfully able to use the ENCODE regulatory network to identify and analyze key TFs in cancer gene expression data.

Page 4: [180] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:01:00 AM

The TF analysis shows that we are successfully able to use the ENCODE regulatory network to identify and analyze key TFs in cancer gene expression data.

Page 4: [180] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:01:00 AM

The TF analysis shows that we are successfully able to use the ENCODE regulatory network to identify and analyze key TFs in cancer gene expression data.

Page 4: [181] Moved from page 4 (Move #4)Microsoft Office User 3/9/17 11:17:00 AM

For example, the ENCODE eCLIP experiment has profiled many SUB1 peaks on genethe ’3’UTR regions of geness, and we find that the predicted targets of the RBP SUB1 were significantly up-regulated in many cancer types (Fig. 4C).

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [182] Deleted Microsoft Office User 3/10/17 1:55:00 PM

Page 4: [183] Moved from page 4 (Move #5)Microsoft Office User 3/9/17 11:17:00 AM

Previously, SUB1 was considered as a TF. However, the ENCODE eCLIP experiment has profiled many SUB1 peaks on gene'3UTR regions. In HepG2 cell where SUB1 eCLIP experiment was done, we found that the decay rate of SUB1 target genes are significantly shorter than non-targets (Fig. 4C). After knocking down SUB1, its predicted targets are also down regulated comparing to other genes (Fig. 4D). These results indicate that SUB1 may bind geneto 3'UTR regions to stabilize transcript slevel. From our integrated analysis, the higher SUB1 activity through regulatory binding on 3'UTR regions is likely to drive tumor-specific expression patterns in many cancer types.

Page 4: [184] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:12:00 AM

Previously, SUB1 was considered as a TF. However, the ENCODE eCLIP experiment has profiled many SUB1 peaks on gene'3UTR regions. In HepG2 cell where SUB1 eCLIP experiment was done

Page 4: [184] Deleted jingzhang.wti.bupt@gmail.com 3/10/17 10:12:00 AM

Previously, SUB1 was considered as a TF. However, the ENCODE eCLIP experiment has profiled many SUB1 peaks on gene'3UTR regions. In HepG2 cell where SUB1 eCLIP experiment was done

Page 4: [185] Deleted Microsoft Office User 3/10/17 4:31:00 PM

gene

Page 4: [185] Deleted Microsoft Office User 3/10/17 4:31:00 PM

gene

Page 4: [185] Deleted Microsoft Office User 3/10/17 4:31:00 PM

gene

Page 4: [185] Deleted Microsoft Office User 3/10/17 4:31:00 PM

gene

Page 4: [186] Moved from page 4 (Move #6)Microsoft Office User 3/9/17 11:17:00 AM

Moreover, the up-regulation of SUB1 target genes is correlated with a worsepoorer patient survival in cancer types such as lung cancer [Office11](Fig. 4).

Page 4: [187] Commented **Microsoft Office User** **3/10/17 4:32:00 PM**

maube change "cancer types such as lung cancer" to just "lung cancer, among other cancer types"?

Page 4: [188] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:02:00 AM**

.

Page 4: [188] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:02:00 AM**

.

Page 4: [189] Moved from page 4 (Move #7)Microsoft Office User **3/9/17 11:17:00 AM**

Step-wise prioritization schemes pinpoint deleterious SNVs in cancer

Page 4: [190] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:19:00 AM**

Not Highlight

Page 4: [191] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:19:00 AM**

No Spacing,TextBody

Page 4: [192] Moved from page 2 (Move #3)Microsoft Office User **3/9/17 11:17:00 AM**

In summary, our cancer encyclopedia consists constitutes a list of regions with higher -than -expected mutational frequencyies in cancer, accurately determined enhancers and gene linkages, the extended gene neighborhoods, thea regulatory network of TFs (and for some lines, a regulatory RBPs network). The cancer encyclopedia also provides , as well as the characteristics of TF/RBPs within the network, such as positions within the network hierarchy, rewiring status, and tumor/normal differential expression in driving potentialcancer, along with and prognostic valuevalues in various cancer types[Office12]. Collectively, these resources allow us to prioritize a few key elements as being associated with oncogenesis, some of which are then validated using small-scale experiments (see table S1).

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [193] Deleted **Microsoft Office User** **3/10/17 4:33:00 PM**

consists

Page 4: [194] Commented **Microsoft Office User** **3/10/17 4:36:00 PM**

You might want to double-check – are my edits correct here?

Page 4: [195] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:19:00 AM**

, some of which are then validated using small-scale experiments (see table S1)

Page 4: [195] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:19:00 AM**

, some of which are then validated using small-scale experiments (see table S1)

Page 4: [196] Moved to page 4 (Move #4) **Microsoft Office User** **3/9/17 11:17:00 AM**

For example, the predicted targets of the RBP SUB1 were significantly up-regulated in many cancer types (Fig. 4C).

Page 4: [196] Moved to page 4 (Move #4) **Microsoft Office User** **3/9/17 11:17:00 AM**

For example, the predicted targets of the RBP SUB1 were significantly up-regulated in many cancer types (Fig. 4C).

Page 4: [197] Moved to page 4 (Move #5) **Microsoft Office User** **3/9/17 11:17:00 AM**

Previously, SUB1 was considered as a TF. However, the ENCODE eCLIP experiment has profiled many SUB1 peaks on gene'3UTR regions. In HepG2 cell where SUB1 eCLIP experiment was done, the decay rate of SUB1 target genes are significantly shorter than non-targets (Fig. 4C). After knocking down SUB1, its predicted targets are also down regulated comparing to other genes (Fig. 4D). These results indicate that SUB1 may bind gene 3'UTR regions to stabilize transcript level. From our integrated analysis, the higher SUB1 activity through regulatory binding on 3'UTR regions is likely to drive tumor specific expression patterns in many cancer types.

Page 4: [198] Moved to page 4 (Move #7) **Microsoft Office User** **3/9/17 11:17:00 AM**

Step-wise prioritization schemes pinpoint deleterious SNVs in cancer

Page 4: [199] Deleted **Microsoft Office User** **3/9/17 11:17:00 AM**

changes

Page 4: [199] Deleted **Microsoft Office User** **3/9/17 11:17:00 AM**

changes

Page 4: [200] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:32:00 AM**

Superscript

Page 4: [201] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:33:00 AM**

Highlight

Page 4: [202] Deleted **Microsoft Office User** **3/10/17 4:39:00 PM**

control

Page 4: [202] Deleted **Microsoft Office User** **3/10/17 4:39:00 PM**

control

Page 4: [203] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:26:00 AM**

on

Page 4: [203] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:26:00 AM**

on

Page 4: [204] Commented **Microsoft Office User** **3/10/17 4:41:00 PM**

Is this written like “CASPER” or “Casper”. Here, you write “CASPER”, but above, this appears as “Casper”

Page 4: [205] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [205] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [205] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [206] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:27:00 AM**

This enhancer is located in a noncoding region. Both histone modification and DHS signals implicate its regulatory role as being active (Fig. 5C), and both our CASPER enhancer prediction method and the STARR-seq experiment support its enhancer function (Fig. 5D). Hi-C and ChIA-PET data link this region to a downstream gene SYCP2. [[what does this mean?]]

Page 4: [207] Deleted **Microsoft Office User** **3/10/17 4:42:00 PM**

the C

Page 4: [207] Deleted **Microsoft Office User** **3/10/17 4:42:00 PM**

the C

Page 4: [207] Deleted **Microsoft Office User** **3/10/17 4:42:00 PM**

the C

Page 4: [208] Formatted	Microsoft Office User	3/10/17 4:44:00 PM

No Spacing,TextBody, Left, Line spacing: single

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [209] Deleted	Microsoft Office User	3/9/17 11:17:00 AM

values

Page 4: [210] Commented	Microsoft Office User	3/10/17 4:45:00 PM

A little bit vague and awkward. Consider revising.

Page 4: [211] Deleted	Microsoft Office User	3/10/17 1:55:00 PM

Page 4: [211] Deleted	Microsoft Office User	3/10/17 1:55:00 PM

Page 4: [211] Deleted	Microsoft Office User	3/10/17 1:55:00 PM

Page 4: [211] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [211] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [211] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [211] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

Page 4: [212] Commented **Microsoft Office User** **3/10/17 4:50:00 PM**

Very unclear – consider revising

Page 4: [213] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

It is challenging because of the v

Page 4: [213] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

It is challenging because of the v

Page 4: [213] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

It is challenging because of the v

Page 4: [213] Deleted **Microsoft Office User** **3/10/17 1:55:00 PM**

It is challenging because of the v

Page 4: [214] Deleted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:41:00 AM**

A key inspiring aspect of our analysis is that by sheer data integration from various assays we demonstrate importance of more accurate non-coding elements and their linkages definitions to interpret both mutation and expression data in cancer. We can straightforwardly envision a path forward through expendingby expanding this framework by integrating additional types of assays and applying to other cancer relevant cell lines and eventually tissues. We also anticipate that by employing more mutation and expression profiles from larger cohorts will definitely benefit our interpretation. However, there are some limitations in this work. One is to appropriately match cancer and normal cell line/tissues for different cancer types. The major leap would be taken is to do this type of work on real cancer tissue. Nevertheless, we are still very encouraged to see that the resource we collected and some of the frameworks we developed are useful to understand some bits of the expression and mutation drive real tissues.[[more]]

Page 4: [215] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Highlight

Page 4: [215] Formatted **Microsoft Office User** **3/9/17 11:17:00 AM**

Highlight

Page 4: [216] Formatted **jingzhang.wti.bupt@gmail.com** **3/10/17 10:33:00 AM**

Font:Times New Roman